Delays for Last-Come First-Served Service and the Busy Period

By JOHN RIORDAN

(Manuscript received November 10, 1960)

For Poisson input to a single server, it is shown that the stationary delay distribution function for last-come first-served service is equal to the distribution function for the busy period (the interval of time during which the server is continuously busy) only for exponential distribution of service time. A similar argument shows that the identity persists for a group of fully accessible servers, each with exponential distribution of service times — a result which has been a curiosity for some time. Finally, the delay distribution for last-come first-served service and a single server with constant service time is derived.

I. INTRODUCTION

Delays for last-come first-served service were first considered by Vaulot[1] for a system with Poisson input to a group of fully accessible servers, each with exponential service time. For this order of service, an arrival which is not served immediately, following the biblical edict, goes to the head of the waiting line. Its consideration has a natural theoretical interest, because, as the opposite of first-come first-served, it seems to be a bound for the gamut of possible service assignments, or at least of those with simple structure. Indeed Vaulot[1] has used it to find the envelope of delay functions for all service assignments.

Very briefly, Vaulot's formulation is as follows. Let $v_n(t)$ be the probability that a waiting demand for service, which at a given moment has just become $n + 1$ in line, waits at least t [$v_n(t)$ is the complement of a distribution function; $v_0(t)$ is the complement of the conditional delay distribution function]. Then, if a is the arrival rate, and b the service rate for each of the c servers, the set of differential recurrence relations for the $v_n(t)$ is

$$v_n'(t) = bcv_{n-1}(t) - (a + bc)v_n(t) + av_{n+1}(t), \qquad n = 0, 1, \cdots .$$

Vaulot's solution of these equations will be given later. Here it is sufficient to notice that the same relations had already appeared in the formulation of the busy period, for the same traffic system, by Palm.[2] The correspondent to $v_n(t)$ is $f_n(t)$, the probability that a busy period (all servers busy), which at a given moment has n waiting customers, continues as a busy period at least t. Since the boundary conditions also agree, $v_n(t) = f_n(t)$, and in particular $v_0(t) = f_0(t)$; that is, the conditional delay distribution function for last-come first-served service is equal to the distribution function of the busy period, for the given system. This is the curious and puzzling result mentioned in the abstract.

The first result of this paper is the proof that, for Poisson input to a single server, the two distributions are alike only for exponential service. For more than one server, it is plausible that the same thing is true, though no easy line of proof seems open. For many servers, each with exponential service time, proof of identity may be given in a way parallel to the single server case. A second result is the determination of the delay distribution for Poisson input to a single server with constant service time; this result belongs in the book with that of Burke[3] on random service for the same system.

II. LAST-COME FIRST-SERVED DELAY FOR POISSON INPUT TO A SINGLE SERVER

For last-come first-served service, an arrival finding the server busy is delayed until all subsequent arrivals finding the server busy or just completing service have been served. The waiting demands at the arrival epoch have no effect on this delay because the new arrival goes to the head of the line.

Following Takács,[4] the delay distribution may be derived as follows. Take the arrival rate as a, the service rate (reciprocal of the average service time) as b, the service time distribution function as $B(t)$, and the delay distribution function as $G^*(t)$ — the star indicating that $G^*(t)$ differs from $G(t)$, the distribution function for the busy period, in the starting epoch of the corresponding intervals.

Note first that the (stationary) distribution function between an arbitrary time epoch in a service interval and the epoch of next service completion is given by†

$$C(t) = b \int_0^t [1 - B(u)]\, du, \tag{1}$$

† The earliest proof of this result known to me is in Palm[5]; it may have been proved much earlier.

so that

$$\gamma(s) = \int_0^\infty e^{-st}\, dC(t) \tag{2}$$
$$= (b/s)[1 - \beta(s)],$$

with

$$\beta(s) = \int_0^\infty e^{-st}\, dB(t).$$

Note that $\gamma(s) = \beta(s)$ if and only if $\beta(s) = b(b + s)^{-1}$, or what is the same thing, $B(t) = 1 - e^{-bt}$.

Now consider the interval between the arrival epoch of a delayed demand and the epoch of the next service completion. The probability that this interval has length $(y, y + dy)$ is $dC(y)$. The probability of n arrivals in this interval, when its length is y, is the Poisson term $e^{-ay}(ay)^n/n!$ If $n = 0$, the delay of the given arrival for last-come first-served service is simply $C(t)$; if $n = 1$, the delay consists of the interval to the first service completion plus the busy interval occasioned by this arrival and *all subsequent arrivals during subsequent service periods*, the distribution function for which is $G(t)$. Hence for $n = 1$ the delay is the convolution of $C(t)$ and $G(t)$. In the same way, for $n = 2$ the delay is the convolution of $C(t)$,$G(t)$ and $G(t)$, and so on.

If $G_n(t)$ is the distribution function for the sum of n variables, each with distribution function $G(t)$, and $G_0(t) = 1$, $G_1(t) = G(t)$, then the conditional delay distribution function is given by

$$G^*(t) = \int_0^t \sum_{n=0} e^{-ay} \frac{(ay)^n}{n!} G_n(t - y)\, dC(y). \tag{3}$$

If

$$\Gamma^*(s) = \int_0^\infty e^{-st}\, dG^*(t)$$

and $\Gamma(s)$ has a similar significance, then the transform of (3) is†

$$\Gamma^*(s) = \gamma[s + a - a\Gamma(s)], \tag{4}$$

with $\gamma(s)$ defined by (2). Note that $\Gamma(s)$ is determined by Takács' equation

$$\Gamma(s) = \beta[s + a - a\Gamma(s)]. \tag{5}$$

† A similar and equivalent result appears in Wishart[6]; note that Wishart considers the unconditional delay distribution function, which is less directly comparable (than the conditional) with the distribution function of the busy period.

Hence $\Gamma^*(s) = \Gamma(s)$ when and only when $\gamma(s) = \beta(s)$. As already noted, this implies $B(t) = 1 - e^{-bt}$, the exponential service distribution.

The common distribution function for exponential service time has been given by Takács,[4] and is obtained as follows. First, since $\beta(s) = b(b + s)^{-1}$, the solution of (5) is

$$\Gamma(s) = \frac{a + b + s - \sqrt{(a + b + s)^2 - 4ab}}{2s}.$$

Next, the inverse of this† is

$$G(t) = \int_0^t \frac{dx}{x\sqrt{\rho}} e^{-(a+b)x} I_1(2x\sqrt{ab}), \qquad \rho = a/b, \tag{6}$$

with $I_1(x)$ the Bessel function of the first kind and imaginary argument. An equivalent expression due to Vaulot,[1] which seems better adapted to numerical work, is

$$1 - G(t) = \frac{2}{\pi} \int_0^\pi \frac{dx}{A} e^{-Abt} \sin^2 x, \qquad A = 1 + \rho - 2\sqrt{\rho} \cos x. \tag{6a}$$

III. DELAYS FOR POISSON INPUT TO MANY EXPONENTIAL SERVERS

The modifications of the argument above for c fully accessible servers, each with exponential distribution of service time (exponential servers, for brevity), are relatively minor. First, in the derivation of the busy-period distribution, a service period in the single-server case is replaced by the interval between the epoch at which the last idle server becomes busy and the epoch at which the first of c busy servers becomes idle. The distribution function of this interval is that of the least of c random variables, each with the same exponential distribution, the service time distribution. If this function is $H(t)$ and the service time distribution is $B(t) = 1 - e^{-bt}$, then

$$1 - H(t) = [1 - B(t)]^c = e^{-bct}. \tag{7}$$

Hence the distribution function $G_c(t)$ is obtained from its single server (exponential service time) correspondent $G(t)$ simply by replacing b by bc.

The last-come first-served delay distribution is proved identical simply by the remark that $H(t)$ is also the distribution of the interval between an arbitrary point in the interval for which $H(t)$ is the distribution function and its termination, because $H(t)$ is of exponential form.

† Ref. 7, pair 556.1, p. 59.

Fig. 1 shows a comparison of conditional delay curves for various orders of service at an occupancy $\rho = a/bc$ of 0.9. The orders are order of arrival, random, and inverse order of arrival (last-come first-served). The abscissae are values of an auxiliary time variable $u = bct$. The ordinates are values of $F(u)$, the probability that the delay of a delayed arrival is at least u. The curve marked "envelope" is an upper bound for all possible orders of service; actually it is a plot of $1 - v(u)$ with

$$v(u) = (1 - \rho) \sum \rho^n v_n(u)$$

and $v_n(u)$ as defined in the introduction.

A detailed study of delays for last-come first-served service and the busy period, with and without defections from waiting, is planned for a later paper.

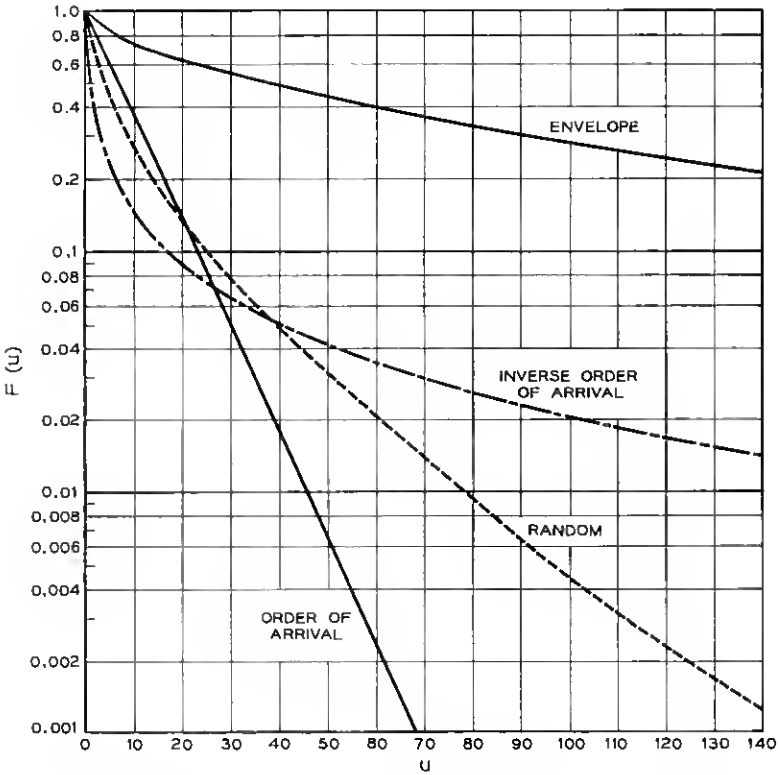

Fig. 1 — Comparison of delay curves for various orders of service occupancy 0.9.

IV. DELAY FOR SINGLE SERVER WITH CONSTANT SERVICE TIME

Although this is a limit case of (4), both delay and busy-period distribution functions are obtained more easily, following Takács for the latter, from the classification of the busy period by number served.

Note first that the generating function for number of arrivals in an interval with distribution function $C(t)$ is given by

$$\begin{aligned} p^*(x) &= \sum p_n^* x^n \\ &= \sum_{n=0} a^n x^n \int_0^\infty e^{-ay}(y^n/n!)\, dC(y) \\ &= \gamma(a - ax). \end{aligned} \tag{8}$$

The corresponding generating function for number of arrivals in a service interval is given by

$$p(x) = \beta(a - ax) \tag{9}$$

and, by (2),

$$p^*(x) = \frac{1 - p(x)}{\rho(1 - x)}, \qquad \rho = a/b. \tag{10}$$

Note that

$$p^*(1) = \frac{p'(1)}{\rho} = \frac{a\beta'(0)}{\rho} = 1,$$

$$\rho p_n^* = 1 - p_0 - p_1 - \cdots - p_n .$$

Now write f_n^* for the probability of n services in a delay period for last-come first-served service, and

$$f^*(x) = \sum_{n=1} f_n^* x^n$$

for its generating function; f_n and $f(x)$ are the corresponding entities for a busy period. Then, following Takács,[4]

$$\begin{aligned} f_1^* &= p_0^*, \\ f_2^* &= p_1^* f_1 , \\ f_3^* &= p_1^* f_2 + p_2^* f_1^2, \end{aligned}$$

and

$$f_n^* = p_1^* f_{n-1} + p_2^* \sum f_j f_{n-1-j} + p_3^* \sum f_j f_k f_{n-1-j-k} + \cdots, \tag{11}$$

$$f^*(x) = xp^*[f(x)] = x\gamma[a - af(x)]. \tag{12}$$

Note that the corresponding result for $f(x)$ obtained by Takács is

$$f(x) = x\beta[a - af(x)]. \tag{13}$$

Hence, by (2), (12), and (13),

$$f^*(x) = \frac{x - f(x)}{\rho[1 - f(x)]}, \qquad \rho = a/b. \tag{14}$$

Note that

$$f'(x) = \frac{f(x)}{x} - axf'(x)\beta'[a - af(x)].$$

Note also that (14) may be rewritten as

$$\begin{aligned} \rho f^*(x) &= 1 + (x - 1)[1 - f(x)]^{-1} \\ &= 1 + (x - 1)g(x), \end{aligned} \tag{14a}$$

where

$$\begin{aligned} g(x)[1 - f(x)] &= 1, \\ g(0) = g_0 &= 1. \end{aligned} \tag{15}$$

Hence

$$\rho f_n^* = g_{n-1} - g_n . \tag{16}$$

For constant service time with service rate b,

$$\begin{aligned} B(t) &= 0, \qquad t < b^{-1}, \\ &= 1, \qquad t > b^{-1} \end{aligned}$$

and $\beta(s) = e^{-s/b}$. Hence

$$f(x) = xe^{-\rho+\rho f(x)},$$

whose solution is given by

$$f(x) = \sum_{n=1} \frac{(\rho n)^{n-1}e^{-n\rho}}{n!} x^n,$$

so that

$$f_n = \frac{(\rho n)^{n-1}e^{-n\rho}}{n!}. \tag{17}$$

Then, using (15), the coefficients g_n of the generating function $g(x)$, are given by

$$g_0 = 1,$$

$$g_1 = f_1 = e^{-\rho},$$

$$g_2 = f_2 + f_1^2 = (1 + \rho)e^{-2\rho},$$

$$g_3 = f_3 + 2f_2f_1 + f_1^3 = (1 + 2\rho + 3\rho^2/2!)e^{-3\rho}.$$

These results suggest writing

$$g_n = g_n(\rho) = \sum_{k=0}^{n-1} \frac{e^{-n\rho} g_{nk} \rho^k}{k!} . \tag{18}$$

From (15) it follows that

$$g_n = \sum_{j=1}^{n} f_j g_{n-j} . \tag{18a}$$

From this and (17) and (18), it is found that

$$g_{nk} = \sum_{j=0}^{k} \binom{k}{j} (j + 1)^{j-1} g_{n-j-1,k-j} . \tag{19}$$

From (19) it is found in succession that $g_{n0} = 1$, $g_{n1} = n - 1$, $g_{n2} = n(n - 2)$, all of which are contained in the single formula $g_{nk} = n^{k-1}(n - k)$. If this is correct, substitution in (19) shows that the following must be an identity:

$$n^{k-1}(n - k) = \sum_{j=0}^{k} \binom{k}{j} (j - 1)^{j-1}(n - j - 1)^{k-j-1}(n - k - 1). \tag{20}$$

Equation (20) is in fact one of the forms associated with Abel's generalization of the binomial formula; it is a special case of Equation (1b) of Salie,[8] which is due to Jensen[9] (of whom it is proper to remember here that he was chief engineer of the Copenhagen Telephone Company as well as a mathematician of note).

Thus, finally,†

$$g_n = e^{-n\rho} \sum_{k=0}^{n} (n - k) n^{k-1} \rho^k / k! \tag{21}$$

which, with (16), determines $f_n^*(x)$.

Note also that the polynomials $g_n \equiv g_n(\rho)$ appear also in the dual

† An equivalent result has been found independently by Frank A. Haight, of the University of California.

problem of the busy period of a single server with exponential service time and regular arrivals considered by Connolly.[10]

The first service has the distribution function

$$\begin{aligned} S_1(t) &= bt, \qquad t \leqq b^{-1}, \\ &= 1, \qquad t \geqq b^{-1} \end{aligned} \tag{22}$$

and the distribution function for n services is

$$\begin{aligned} S_n(t) &= 0, \qquad t \leqq (n-1)b^{-1}, \\ &= bt, \qquad (n-1)b^{-1} < t < nb^{-1}, \\ &= 1, \qquad nb^{-1} < t. \end{aligned} \tag{23}$$

So

$$G^*(t) = \sum_{n=1} f_n^* S_n(t) \tag{24}$$

is the (conditional) delay distribution function for last-come first-served service with Poisson input to a single server with constant service time. The distribution function for the busy period (Takács[4]) is

$$G(t) = \sum_{n=0}^{[bt]} e^{-\rho n}(\rho n)^{n-1}/n!, \tag{25}$$

with $[bt]$ the integral part of bt.

REFERENCES

1. Vaulot, E., Delais d'attente des appels téléphoniques dans l'ordre inverse de leur arrivée, C. R. Acad. Sci. (Paris), **238**, 1954, pp. 1188–1189.
2. Palm, C., Specialnummer för teletrafikteknik, Tekniska Meddelanden från Kungl. Telegrafstyrelsen, 1946; English version: Research on Telephone Traffic Carried by Full Availability Groups, Tele, No. 1, 1957.
3. Burke, P. J., Equilibrium Delay Distribution for One Channel with Constant Holding Time, Poisson Input and Random Service, B.S.T.J., **38**, 1959, pp. 1021–1031.
4. Takács, L., Investigation of Waiting Time Problems by Reduction to Markov Processes, Acta Math. Acad. Sci. Hung., **6**, 1955, pp. 101–129.
5. Palm, C., Intensitätsschwankungen im Fernsprechverkehr, Ericsson Tech., No. 44, 1943.
6. Wishart, D. M. G., Queuing Systems in Which the Discipline is "Last-Come, First-Served", Oper. Res., **8**, 1960, pp. 591–599.
7. Campbell, G. A., and Foster, R. M., *Fourier Integrals for Practical Applications*, D. Van Nostrand Co., New York, 1948.
8. Salie, H., Über Abel's Verallgemeinerung der binomischen Formel, Ber. Verh. Sächs. Akad. Wiss. Leipzig Math.-Nat. Kl., **98**, 1951, pp. 19–22.
9. Jensen, J. L. W. V., Sur une identite d'Abel et sur d'autres formules analogures, Acta Math. **26**, 1902, pp. 307–318.
10. Connolly, B. W., The Busy Period in Relation to the Queuing Process GI/M/l, Biometrika, **46**, 1959, pp. 246–251.